Probability for Stat 343

Notation

We will use capital letters like X and Y to represent random variables, and lower case x and y to denote observed values or values we might hypothetically observe. I will sometimes also use capital letters to stand for matrices, and we'll just have to be clear from the context about what is a matrix and what is a random variable.

In type written text, I will use bold letters to denote vectors, which are column vectors by default:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} \text{ is a random vector and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \text{ is a vector of observed values.}$$

When writing on the board I will use a squiggly underline to denote vectors:

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} \text{ is a random vector and } \underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \text{ is a vector of observed values.}$$

Probability Mass/Density Function (p.m.f., p.d.f.), Cumulative Distribution Function (c.d.f.):

If X is a discrete random variable, then the probability mass function $f_X(x) = P(X = x)$

If X is a continuous random variable, then the probability density function $f_X(x)$ can be used to find $P(X \in [a, b]) = \int_a^b f_X(x)dx$

The cumulative distribution function is $F(x) = P(X \leq x)$:

- If X is discrete, $F(x) = \sum_{t=-\infty}^{x} f_X(t)$
- If X is continuous, $F(x) = \int_{-\infty}^{x} f_X(t)dt$

Joint Distributions from the Marginals

If X and Y and both discrete, then they have a joint p.m.f.: $f_{X,Y}(x, y) = P(X = x \text{ and } Y = y)$

If X and Y are both continuous, then they have a joint p.d.f.: $P(X \in [a, b] \text{ and } Y \in [c, d]) = \int_a^b \int_c^d f_{X,Y}(x, y)\, dx\, dy$

If one of X and Y is discrete and the other is continuous, it's possible to define a similar probability function $f_{X,Y}(x, y)$.

If X and Y are independent, then their joint p.f. is the product of their marginal p.f.'s:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

If X and Y are **not** independent, their joint p.f. is the product of the marginal for one and the conditional for the second given the first:

$$f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x) = f_Y(y)f_{X|Y}(x|y)$$

Marginal distributions from the Joint

Suppose X and Y have joint probability function $f_{X,Y}(x, y)$.

If X is discrete, then the marginal probability function for Y is $f_Y(y) = \sum_x f_{X,Y}(x, y)$

If X is continuous, then the marginal probability function for Y is $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$

Conditional Distributions

By definition, the conditional distribution for $Y|X = x$ has p.f. $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.

This also extends to more random variables. For example $(W, X)|Y = y, Z = z$ have the joint conditional distribution with p.f. $f_{W,X|Y,Z}(w, x|y, z) = \frac{f(w,x,y,z)}{f(y,z)}$

Bayes' Rule

If I know the marginal distribution for X has p.f. $f_X(x)$ and the conditional distribution for $Y|X$ has p.f. $f_{Y|X}(y|x)$ then I can calculate the p.f. for the conditional distribution of $X|Y$ as follows:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y)dx} \\ &= \frac{f_X(x)f_{Y|X}(y|x)}{\int f_X(x)f_{Y|X}(y|x)dx} \end{aligned}$$

If X is discrete, replace the integral in the denominator by a summation.

There are two ways of explaining why this is useful:

1. It lets us reverse the order of conditioning from $Y|X$ (what we know to start with) to $X|Y$.
2. It lets us update our knowledge about the distribution of X having observed a value of $Y = y$.

Expected Value and Variance

$$E(X) = \int x f_X(x)dx$$

$$\begin{aligned} Var(X) &= \int (x - E(X))^2 f_X(x)dx \\ &= \int (x^2 - 2xE(X) + E(X)^2) f_X(x)dx \\ &= \int x^2 f_X(x)dx - 2E(X) \int xdx + E(X)^2 \int f_X(x)dx \\ &= E(X^2) - E(X)^2 \end{aligned}$$

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2 Var(X)$$

Central Limit Theorem

There are many slightly different statements of the central limit theorem. Here's one:

Formal statement in a not-too-useful form

Suppose $X_1, X_2, \ldots$ is a sequence of independent, identically distributed (i.i.d.) random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Define the sequence of random variables $Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu\right)$. Then as n approaches infinity, the random variables Z_n converge in distribution to a Normal$(0, 1)$ random variable.

Informal statement, still not too useful

If n is "large enough" (how large? it depends.), it's approximately true that

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu\right) \sim \text{Normal}(0, 1)$$

as long as the X_i are i.i.d. with mean μ and variance σ^2

Some intermediate steps

Let's multiply Z_n by $\frac{\sigma}{\sqrt{n}}$ and add μ.

Since $Z_n \sim \text{Normal}(0, 1)$ (approximately, for large n), $\frac{\sigma}{\sqrt{n}} Z_n + \mu \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$ (approximately, for large n).

We see that

$$\begin{aligned}
\frac{\sigma}{\sqrt{n}} Z_n + \mu &= \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma}\left(\frac{\sum_{i=1}^n X_i}{n} - \mu\right) + \mu \\
&= \frac{\sum_{i=1}^n X_i}{n} - \mu + \mu \\
&= \frac{\sum_{i=1}^n X_i}{n}
\end{aligned}$$

Informal statement, more useful

If n is "large enough" (how large? it depends.), it's approximately true that

$$\frac{\sum_{i=1}^n X_i}{n} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

as long as the X_i are i.i.d. with mean μ and variance σ^2